

AD-A282 411



MENTATION PAGE

Form Approved
OMB No 0704-0188

Do not write in this space. It is reserved for the use of the reviewing instructions, searching and data sources, and for the preparation of information. Send comments regarding this burden estimate or any other aspect of this form, including suggestions for reducing the burden, to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

REPORT DATE

3. REPORT TYPE AND DATES COVERED

THESIS/DISSERTATION

4. TITLE AND SUBTITLE

Expression and Partial Purification
of Several Truncated Forms of Chloroplast
Translational Initiation Factor 2 From
Euglena gracilis

5. FUNDING NUMBERS

6. AUTHOR(S)

Kevin E. Blanchard

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

AFIT Student Attending:

Univ. of North Carolina

8. PERFORMING ORGANIZATION
REPORT NUMBER

AFIT/CI/CIA-

94053

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

DEPARTMENT OF THE AIR FORCE

AFIT/CI

2950 P STREET

WRIGHT-PATTERSON AFB OH 45433-7765

10. SPONSORING/MONITORING
AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for Public Release IAW 190-1
Distribution Unlimited
MICHAEL M. BRICKER, SMSgt, USAF
Chief Administration

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

DTIC
S ELECTE
JUL 21 1984
F9818 94-22720

DTIC QUALITY INSPECTED 8

94 7 19 197

14. SUBJECT TERMS

15. NUMBER OF PAGES

57

16. PRICE CODE

17. SECURITY CLASSIFICATION
OF REPORT18. SECURITY CLASSIFICATION
OF THIS PAGE19. SECURITY CLASSIFICATION
OF ABSTRACT

20. LIMITATION OF ABSTRACT

AFIT/G-94-053

Expression and Partial Purification of
Several Truncated Forms of Chloroplast
Translational Initiation Factor 2 from
Euglena gracilis

By

Kevin E. Blanchard

1994

Advisor, Professor Linda L. Spremulli

Accession For	
NTIS CRA&I	<input checked="checked" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

This report is submitted to fulfill the written requirements
for the Master of Science degree in the Department of
Chemistry at the University of North Carolina at Chapel
Hill.

ABBREVIATIONS

A ₂₈₀	absorbance at 280 nanometers
ATP	adenosine triphosphate
βMe	2-mercaptoethanol
DEAE	diethylaminoethyl
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	ethylenediaminetetraacetic acid
<i>E. gracilis</i>	<i>Euglena gracilis</i>
eIF-2	eukaryotic translational initiation factor 2
fMet-tRNA ^{Met} _f	formylated methionyl-tRNA
HEPES	4-(2-hydroxyethyl)-1-piperazine ethanesulfonic acid
HPLC	high-pressure liquid chromatography
IF-2	translational initiation factor 2
IF-2 _{chl}	chloroplast translational initiation factor 2
IPTG	isopropyl β-D-thiogalactopyranoside
kb	kilobase
kDa	kilodalton
LB	Luria broth
MES	2-(N-morpholino)-ethanesulfonic acid
NaI	sodium iodide
Ni-NTA	nickel nitrilotriacetic acid
PCR	polymerase chain reaction

SDS-PAGE

sodium dodecyl sulfate polyacrylamide gel
electrophoresis

Taq

Thermus aquaticus

ACKNOWLEDGEMENTS

I would like to thank all of the members of the Spremulli lab whose help and support made this work possible. Especially, I would like to say thanks to Dr. Lisa Benkowski and Dr. Mary Farwell for lending their protein purification expertise; to Qiong Lin for the expression plasmids and answering thousands of questions; and to Jim Ward for his willing extra set of hands, knowledge of "all things computer," and good humor in the lab.

Thanks also goes to Dr. Lan Ma for her unlimited patience while teaching me three years worth of laboratory techniques in three months, as well as providing priceless insight on life outside of science.

Finally, my heartfelt thank you to Dr. Linda Spremulli for her unlimited knowledge, infinite patience, and caring enough to always give the time she did not have. Her understanding made finishing this degree on time a reality.

My gratitude also goes to the Air Force Institute of Technology for providing the scholarship that allowed me to pursue a graduate degree.

TABLE OF CONTENTS

ABBREVIATIONS.....	I
ACKNOWLEDGEMENTS.....	III
TABLE OF CONTENTS.....	IV
ABSTRACT.....	VI
INTRODUCTION.....	1
Figure 1.....	4
MATERIALS AND METHODS.....	8
Materials.....	8
Figure 2.....	10
Buffers.....	11
PCR Reactions.....	11
Ligation/Transformation.....	11
Figure 3.....	13
Small-Scale Expression.....	15
Purification of the G-Domain.....	15
Figure 4.....	18
Miscellaneous Procedures.....	19
RESULTS.....	20
Construction of Expression Plasmids.....	20
Confirmation of The Sequences of the Expression Constructs.....	21
Figure 5.....	23
Figure 6.....	25

Figure 7.....	27
Figure 8.....	29
Growth Rate of Strains Containing Expression Plasmids.....	30
Small-Scale Expression of the G-Domain Protein.....	30
Figure 9.....	32
Figure 10.....	34
Figure 11.....	36
Small-Scale Expression of the IF-2 _{chl} Gamma Protein.....	37
Purification of the G-Domain.....	37
Figure 12.....	39
Figure 13.....	42
Figure 14.....	44
Figure 15.....	46
Figure 16.....	48
Figure 17.....	50
DISCUSSION.....	52
REFERENCES.....	56
APPENDIX A	
APPENDIX B	

ABSTRACT

Two truncated forms of *Euglena gracilis* chloroplast initiation factor 2 have been engineered at the DNA level and expressed at a high level in *Escherichia coli*. The DNA was obtained by PCR amplification of a cDNA encoding a portion of IF-2_{chl}. The PCR primers were custom-designed and added a restriction site at each end for sub-cloning into the QIAexpress expression system. This expression system provides excellent protein yield in *E. coli* and adds six histidine residues which bind Ni-NTA affinity resin.

Two polypeptides were produced. A 30 kDa polypeptide which contains the highly homologous guanine nucleotide binding region of IF-2_{chl} was designed and has been designated the G-domain. The second protein, designated IF-2_{chl} gamma, has been designed to be approximately the same size (66 kDa) as *E. coli* IF-2 gamma. Based on sodium dodecyl sulfate polyacrylamide gel electrophoresis analysis, both proteins have been expressed in *E. coli*, and production of the G-domain was superb. IF-2_{chl} gamma was present at a much lower relative concentration and its presence greatly retarded cell growth.

The G-domain protein was also purified under native conditions to approximately 85-90 % purity. A four step process utilizing affinity chromatography and high-pressure

liquid chromatography was followed, and the presence of the protein was monitored by SDS-PAGE.

INTRODUCTION

The two energy producing organelles of eukaryotic cells, chloroplasts and mitochondria, each possesses its own genome. They also contain the necessary machinery to transcribe and translate this genetic information into RNAs and proteins (1). Previously in this laboratory, initiation factor 2 (IF-2_{chl}), one of the protein factors for initiation of translation in the chloroplast of *E. gracilis*, was purified and characterized (2). IF-2_{chl} is encoded by a nuclear gene and a portion of its cDNA has been cloned (3). In this report, the sub-cloning and expression of various truncated forms of IF-2_{chl} is reported.

Relatively little is understood about IF-2_{chl}, but IF-2 from the prokaryote, *Escherichia coli*, has been extensively studied and characterized. *E. coli* IF-2 was used as a model for the study of IF-2_{chl} for several reasons. First, *E. coli* IF-2 and *E. gracilis* IF-2_{chl} share significant homology in the C-terminal region (Appendix A). Second, chloroplasts are believed to have arisen in evolution through an endosymbiotic relationship between an early cyanobacterium and an early eukaryotic cell (1). Thus, though *E. gracilis* is a eukaryote, one would expect the translational system in the chloroplast to more closely resemble those in prokaryotic organisms. Third, the eukaryotic homolog to IF-2, eIF-2, is a heterotrimeric

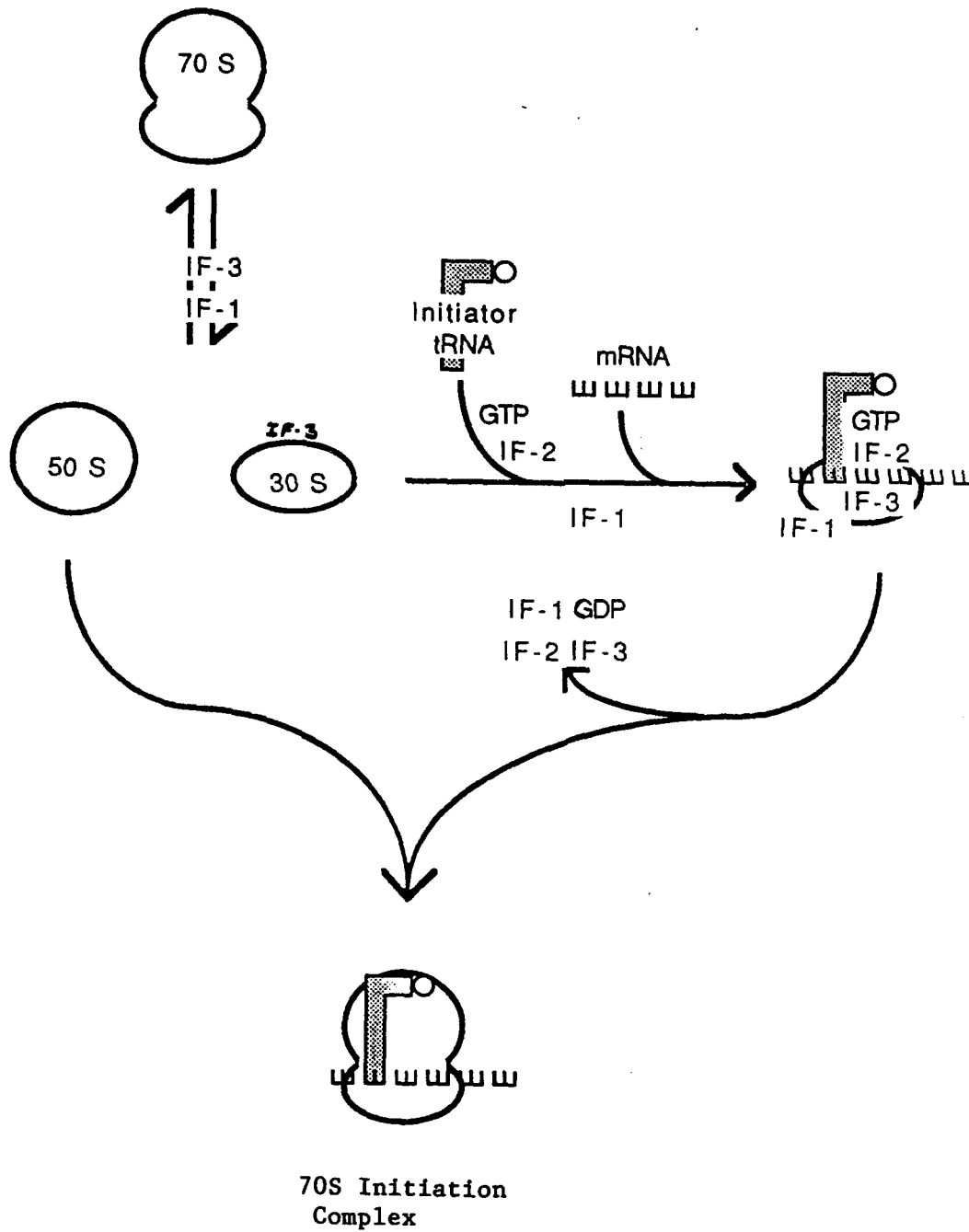
protein that much less closely resembles IF-2_{chl} than does prokaryotic IF-2. Therefore, IF-2 from *E. coli* makes an excellent model for this investigation.

E. coli IF-2 can be found in two forms *in vivo*, 97.3 kDa and 79.7 kDa. These two forms arise as a result of the presence of two translational start sites on the mRNA (4). The two forms are designated alpha and beta, respectively (4). Another form, gamma (65.4 kDa), can be isolated and appears to be the result of proteolytic cleavage during purification (4). All three forms differ only in their N-terminal ends; and all are equally active, though gamma exhibits slightly decreased stability (5). Furthermore, Laalami et al. (6) report that a genetically engineered 55 kDa C-terminal fragment of *E. coli* IF-2 is active *in vivo*.

IF-2 has a major role in the initiation of translation in prokaryotes. (See Figure 1 for a schematic overview of initiation in prokaryotes.) This factor is essential for the formation of the translational initiation complex because it facilitates the binding of the initiator tRNA (fMet-tRNA_f^{Met}) to the 30S ribosomal subunit in the presence of mRNA (4). GTP is also required for the formation of the initiation complex and is bound by IF-2. IF-2 acts as a ribosome dependent GTPase upon the binding of the 50S ribosomal subunit to the initiation complex (7). These functions have been qualitatively assigned to certain regions in IF-2.

Figure 1: Schematic overview of initiation of translation in *E.coli* (8).

General Scheme for the Initiation of
Protein Biosynthesis in E. coli



Based on homology to other G-proteins and GDP binding studies on engineered forms of IF-2, the GTP binding function has been assigned to amino acids 392-540 of the 890 amino acid alpha form of IF-2 (4). Also, based on the ability of the 55 kDa protein to sustain growth in *E. coli* when supplied in excess, the fMet-tRNA_f^{Met} binding function has been assigned downstream of residue 540. Thus, the first 42.8 kDa of the alpha form is not required for cell viability when the C-terminal portion of the factor is available in elevated amounts. (4).

With this understanding of *E. coli* IF-2 as a frame of reference, it is appropriate to examine what is known about IF-2_{chl} from *E. gracilis*. IF-2_{chl} has been isolated in two main forms, designated alpha and beta, based on their affinity for DEAE-cellulose (2). The alpha form can be divided further into alpha-I and alpha-II. The alpha-I forms of IF-2_{chl} have molecular masses ranging from 700-800 kDa and are composed of tetramers of 200 kDa subunits. The alpha-II form of IF-2_{chl} has a molecular mass in the range of 200-400 kDa and is a multimer of 97 and 110 kDa subunits. The beta form has a molecular mass of 200 kDa and is composed of 97 kDa subunits (2). Though IF-2_{chl} is vastly different in size and in its state of aggregation from *E. coli* IF-2, its biological function is quite similar. All three forms of IF-2_{chl}, like *E. coli* IF-2, promote fMet-tRNA_f^{Met} binding to chloroplast 30S ribosomal subunits

in a message-dependent manner (2). As in prokaryotic initiation, this step requires GTP (2). IF-2_{chl} is not active on *E. coli* ribosomes; but IF-2 from *E. coli* is active with chloroplast ribosomes (2). Finally, it is not known if the 97, 110, or 200 kDa subunits of the various forms of IF-2_{chl} are active as monomers or how they originate. Based on the presence of intermediate sizes of the protein in sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and Western blot experiments, as well as mRNA analysis, (3) it is believed that this protein is probably made as a 200 kDa monomer and proteolytically processed in the cell (2).

Part of the cDNA derived from the nuclear-encoded gene for IF-2_{chl} has been cloned and sequenced in this laboratory resulting in several interesting facts. First, a Northern blot showed an mRNA approximately 6.5 kb in length (3) which could correspond to a protein over 200 kDa in size. Using a random-primed cDNA library 2850 base pairs of the cDNA have been cloned and sequenced from the probable center of the gene to the poly(A) tail (9) including a 2466 nucleotide open reading frame (Appendix B). The translated protein sequence of the region of IF-2_{chl} characterized to date is homologous to *E. coli* IF-2, especially in the C-terminal half of the protein. The overall identity is 38 % (Appendix A). IF-2_{chl} is 75 percent identical to *E. coli* IF-2 in the G-domain region (residues 392-540; Appendix A).

Two polypeptides, corresponding to different regions of IF-2_{chl}, have been engineered at the DNA level and expressed in *E. coli*. One polypeptide represents the G-domain and some surrounding residues of IF-2_{chl}. The second is a protein designed to be the same size as the *E. coli* gamma form of IF-2. This protein will be referred to as IF-2_{chl} gamma. The partial purification of the G-domain protein has also been accomplished.

MATERIALS AND METHODS

Materials: Taq DNA polymerase, T4 ligase, restriction enzymes and appropriate reaction buffers were obtained from Promega. The SeqLEnd oligonucleotide was synthesized by the Pathology Department at the University of North Carolina at Chapel Hill (UNC-CH) and the oligonucleotides 5-gamma and 3-gdomain were synthesized in the Lineberger Comprehensive Cancer Center at the UNC-CH (Figure 2). Kanamycin, ampicillin, isopropyl β -D-thiogalactopyranoside (IPTG), alumina A305, ethidium bromide, and 2-(N-morpholino)-ethanesulfonic acid (MES) were purchased from Sigma. Nickel nitrilotriacetic acid (Ni-NTA) resin, *E. coli* M15, and pQE vectors were obtained from QIAGEN as part of a QIAexpress kit. A GENECLAN kit for DNA isolation was purchased from BIO 101, and SDS-PAGE pre-stained molecular weight markers as well as reagents for protein determination by the Bradford method were obtained from Bio-Rad. Tris, β -mercaptoethanol (β Me), ethylenediaminetetraacetic acid (EDTA), and agarose were purchased from Fisher and HEPES from Mallinckrodt. High-pressure liquid chromatography (HPLC) columns TSKgel DEAE-5PW (7.5 mm x 75 mm) and SP-5PW (7.5 mm x 75 mm) were purchased from Beckman. IF-2_{chl} clones in a pBluescript plasmid were kindly provided by Dr. Lan Ma who had done the preliminary cloning work.

Figure 2: PCR Primers. The nucleotide sequences of the three oligonucleotide primers used in the PCR reactions are shown. The primers 5-gamma and 3-gdomain were used for the G-domain construct. The gamma construct was synthesized with 5-gamma and SeqLEnd. Bases in *italics* are recognition sites for restriction enzymes. Bases in **bold** lettering base pair with bases in the template DNA.

Nucleotide Sequences of Oligonucleotide Primers

Name	Sequence
5-gamma	5' GACGGTACCTAGGGATGACCGTCAGCGAG 3'
3-gdomain	5' GACAGATCTAGGTTGGCCTTCAGGTCCGC 3'
SeqLEnd	5' GCCAAGATCTGCGGGGGTTCGCCTCGAGGG 3'

Buffers: Buffer I consisted of 20 mM Tris-HCl pH 7.8, 50 mM NH₄Cl, 10 mM MgCl₂, 6 mM β Me, and 10 % glycerol. Buffer II contained 20 mM HEPES-KOH, pH 7.6, 50 mM NH₄Cl, 0.1 mM EDTA, 6 mM β Me, and 10 % glycerol. Buffer III was identical to Buffer II except that it contained 300 mM NH₄Cl. Buffer IV contained 20 mM MES-KOH, pH 6.0, 50 mM NH₄Cl, 0.1 mM EDTA, 6 mM β Me, and 10 % glycerol. Buffer V was similar to Buffer IV except that it contained 300 mM NH₄Cl.

PCR Reactions: Plasmid DNA for the polymerase chain reaction (PCR) template was a pBluescript vector, designated T5, containing part of the IF-2_{chl} cDNA clone from residue 580 to base 2466 (Appendix B). PCRs were carried out in 0.1 ml reaction mixtures with 2.5 units of Taq DNA polymerase, 50 pmol of each oligonucleotide primer (Figure 2), and approximately 50 fmol of plasmid DNA. The reactions were initiated after a 5 minute denaturation at 85 °C at a MgCl₂ concentration of 1.5 mM. A Perkin Elmer Cetus thermal cycler was used for the reactions and programmed as shown in Figure 3. PCR results were analyzed on 1 % agarose gels and stained with ethidium bromide.

Ligation/Transformation: Aliquots (12 - 20 ul) of the PCR reaction mixtures were removed and added to 0.3 ml of NaI solution (GENECLEAN kit). The PCR product was isolated

Figure 3: Program for PCR. The thermal cycler was programmed as shown for PCR. Steps 3 through 5 were repeated for 30 cycles. Annealing temperatures for the G-domain and gamma constructs were 50 and 60 °C. respectively.

Program for PCR Reactions

<u>Step</u>	<u>Time (min)</u>	<u>Temp (°C)</u>
1	5	85
2	1	94
3 (denature)	0.75	94
4 (anneal)	1	50 or 60
5 (extend)	2.5	72
6	5	72
7	N.A.	4

using the GENECLAN method based on the manufacturer's instructions. DNA (1 ug) was digested with BglII and KpnI and the DNA fragments present were separated by electrophoresis on a 1 % agarose gel in 40 mM Tris acetate with 10 mM EDTA (TAE). Bands containing the desired DNA were excised from the gel and GENECLANed.

A pQE-52/17 Type III expression construct (10) was kindly provided by Qiong Lin. It contained one of her inserts cloned between the polylinker (located on the pQE-52 portion of the plasmid) and the BglII site (located on the pQE-17 portion of the plasmid) (10). The plasmid was digested with BglII and KpnI (there is a KpnI recognition site in the polylinker), and the DNA fragments were separated by electrophoresis on a 1 % agarose gel in TAE. The vector fragment was excised and eluted using GENECLAN. For the ligation reaction, a 10 ul reaction mixture was prepared and contained about 1 pmol of insert and 0.5 pmol of vector. Incubation was carried out with 4 units of T4 DNA ligase and 1 mM ATP at 16 °C for 14 hours.

E. coli M15 carrying the pREP4 plasmid was transformed by high efficiency electro-transformation on a Bio Rad Gene Pulser with 3 ul of the ligation reaction according to the manufacturer's instructions. After incubation in SOC media (LB with 2.5 mM KCl, 10 mM MgSO₄, and 20 mM glucose) at 37 °C for 1 hour, the cell suspension was plated on LB-agar containing 100 ug/ml ampicillin and 25 ug/ml kanamycin.

Colonies were picked and plated on a master plate. Colonies were screened for plasmids of interest (based upon the size of the insert), and frozen cell stocks were made of those strains carrying plasmids of interest. Plasmids were isolated and sequenced by the Sanger dideoxy method (11) to insure the correct reading frame was maintained in the construct.

Small-Scale Expression: A time course for the expression of various constructs was carried out basically as described by QIAGEN (10) using 1 mM IPTG for induction. A 20 ml culture was used and 1.5 ml of the cells were removed at 0.25, 0.5, 1, 2, 3, and 4 hours for harvesting and freezing. The samples were analyzed by 12 % SDS-PAGE.

Purification of the G-Domain: One liter of *E. coli* M15 carrying the construct of interest was grown as described (10) and induced with 1 mM IPTG. Cells were harvested after 4 hours of induction and collected by centrifugation in a Sorvall H6000A rotor at 4500 rpm and 4 °C. The cell pellet was weighed, ground with 2 times the cell weight of alumina A305, and washed into centrifuge tubes with 2-3 volumes of Buffer I. The lysate was subjected to centrifugation at 10,000 rpm at 4 °C in a Sorvall SS-34 rotor for 10 minutes. The supernatant (S30) was transferred to a fresh tube and subjected to centrifugation at 15,000 rpm for 30 minutes in

a Sorvall SS-34 rotor at 4 °C.

The S30 was stirred with 4 ml of Ni-NTA resin previously equilibrated with Buffer I for 1 hour at 4 °C. The slurry was poured into a column and washed with Buffer I until the A_{280} decreased to baseline. The column was then washed with Buffer I which had been adjusted to pH 6.0 by addition of HCl until the A_{280} decreased to baseline. Protein retained by the resin was eluted using a 20 ml linear gradient from 0 to 0.5 M imadazole in Buffer I. Fractions of 2 ml were collected at a flow rate of 0.5 ml/min. Fractions were analyzed by SDS-PAGE and those containing the G-domain protein were pooled.

Pooled fractions were dialyzed against Buffer II and applied to a TSKgel DEAE-5PW HPLC column previously equilibrated with Buffer II. The column was washed at 1 ml/min until the A_{280} returned to baseline. The column was developed with a linear gradient (30 ml) of 0 to 100 % Buffer III at a flow rate of 0.5 ml/min while collecting 1 ml fractions. Fractions were analyzed and pooled as described above. The pooled fractions were buffer-exchanged with Buffer II using a Centricon 10 concentrator. The protein was again applied to the column above and washed as described. The column was developed with the profile shown in Figure 4, and 1.5 minute fractions were collected. Fractions were analyzed and pooled as described previously. The pooled fractions were dialyzed against Buffer IV and

Figure 4: Elution Gradient Profile for 2nd DEAE HPLC Column.

% Buffer III	Time (min)	Flow Rate (ml/min)
0-40	30	0.5
40	15	0.5
40-80	100	0.3
80	15	0.5
80-100	30	0.5

applied to a TSKgel SP-5PW HPLC column previously equilibrated with Buffer IV at a flow rate of 1 ml/min. The column was washed with Buffer IV until the A_{280} reached baseline. The protein was eluted with a linear gradient (60 ml) of 0 to 100 % Buffer V, and 0.5 ml fractions were collected. Fractions were analyzed by SDS-PAGE as described above. Fractions were fast-frozen at -70 °C using a 2-propanol/dry ice bath between each of the above steps.

Miscellaneous procedures: The Laemmli procedure for SDS-PAGE was followed using 12 % polyacrylamide resolving gels (11). Gels were stained with Coomassie Brilliant Blue. Bradford assays were carried out according to reagent manufacturer's instructions with bovine serum albumin as the standard.

RESULTS

Construction of Expression Plasmids: As mentioned, the pQE-52/17 expression vector had already been constructed in our laboratory. This expression system provides a high level of expression and incorporates a 6 residue histidine tag on the C-terminal end of the polypeptide to aid in purification (10). In order to simplify the construction of the expression plasmids PCR primers were designed with three criteria. First, they had to yield a product that could be ligated into pQE-52/17 and maintain the proper reading frame. Second, primers from the 5' end of the IF-2_{chl} sequence had to have a KpnI site engineered on the 5' end of the primer. Similarly primers from the 3' end of the IF-2_{chl} sequence had to have a BglII site engineered on the 5' end of the primer (Figure 2). Third, the primers had to be designed such that they would provide products coding for IF-2_{chl} G-domain and IF-2_{chl} gamma. The resulting primers are shown in Figure 2 and functioned properly as described below.

The G-domain insert was made using 5-gamma and 3-gdomain primers. The 5-gamma primer defines the 5' end of both constructs, and when a lineup based upon homology of *E. coli* IF-2 and IF-2_{chl} is examined, the primer corresponds to the IF-2_{chl} sequence that is homologous to the start of *E.*

coli IF-2 gamma. The 3-gdomain primer corresponds to the 3' end of the guanine nucleotide binding domain. IF-2_{chl} gamma insert was made using 5-gamma and SeqLEnd primers. SeqLEnd base pairs in the region the 5' side of the stop codon in the 2466 residue open reading frame for IF-2_{chl}.

The polymerase chain reaction for the G-domain produced approximately 8-10 ug of the expected 780 base pair product with no evidence of side reactions (Figure 5). Amplifying the DNA necessary for the IF-2_{chl} gamma product proved more difficult. Attempts were made at several annealing temperatures from 50 - 65 °C. before the results shown were obtained (Figure 6). The reaction produced 8-10 ug of the expected 1762 base pair product, and there was no evidence of spurrious products (Figure 6).

Confirmation of the Sequences of the Expression Constructs:

The plasmids containing the G-domain and gamma constructs were sequenced as described in Materials and Methods (results not shown), using custom designed oligonucleotide primers originally used to sequence the cDNA for IF-2_{chl} (9). The sequences of the inserts matched prior sequences (Appendix B), and the ligated regions were in-frame with the start site provided by the pQE vector (10). Complete sequences of the expressed proteins, based on the DNA sequence, are shown in Figure 7 and Figure 8. The sequences represent the entire expressed protein, including any

Figure 5: Electrophoretic Analysis of PCR Amplification of G-domain Sequences. The photograph of an ethidium bromide stained agarose gel shows the results of the PCR used to produce the DNA for the G-domain construct. Lane 1 - 200 ng of high molecular weight standards (lambda DNA digested with HindIII). Lane 2 - 200 ng of low molecular weight standards (pEC1009 digested with Sau3A). Lane 3 - Lane left blank. Lane 4 - A 10 uL aliquot taken from the PCR reaction tube for the G-domain construct. The PCR reaction was carried out as described in Materials and Methods.

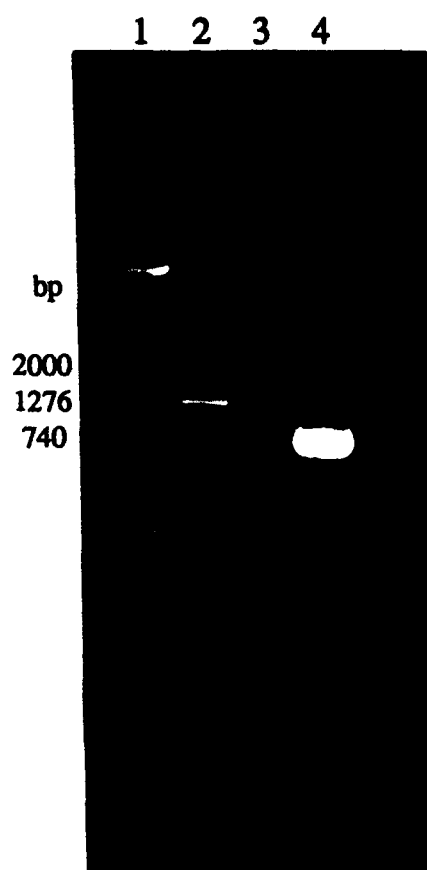


Figure 6: Electrophoretic Analysis of PCR Amplification of IF-2_{chl} Gamma Sequences. The photograph of an ethidium bromide stained agarose gel shows results of the PCR used to produce the DNA for the gamma construct. Lane 1 - 200 ng of high molecular weight markers (lambda DNA digested with HindIII). Lane 2 - 200 ng of low molecular weight markers (pEZC1009 digested with Sau3A). Lane 3 - A 10 uL aliquot taken from PCR reaction tube 1 for the gamma construct. Lane 4 - A 10 uL aliquot taken from PCR reaction tube 2 for the gamma construct. The PCR reaction was carried out as described in Materials and Methods.

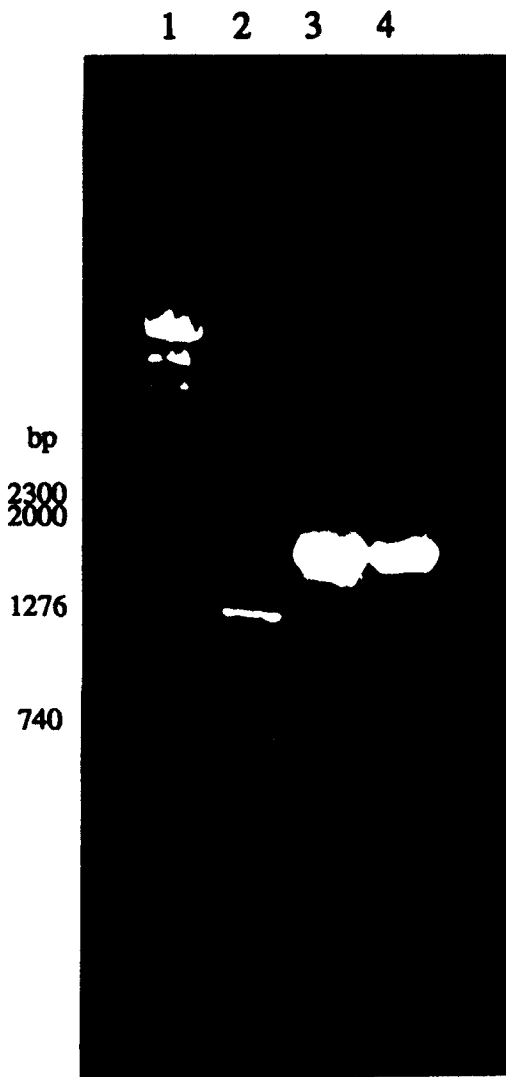


Figure 7: Sequence of G-Domain Protein. The amino acid sequence for the entire G-domain protein as it is expressed is depicted in single letter code. Residues coded by IF-2_{chl} DNA are underlined.

1 MRIRMAARYL GMTVSEIAGK LAITPANVVT VLFKKGIMSA PSOTIAYDLV
51 KIVCDEYKVE VLEVVEEDGI ASMEDRFVLD EEAEALVSRP PVVTIMGHVD
101 HGKTSLLDYI RKSNVVAGEA SGITOAIGAY HVEFASPTDG TPTFISFIDT
151 PGHEAFTAMR ARGATVTDIT IIVVAADDGV RPOTKEAIAH CKAAGVPMVV
201 AINKIDKDGA DPERVMNELA OAGLVPEEWG GEVPTVKISA KKGLGIKELL
251 EMILLTAEVA DLKANLDLRS HHHHHH*

Figure 8: Sequence of IF-2_{chl} Gamma Protein. The amino acid sequence for the entire IF-2_{chl} gamma protein as it is expressed is depicted in single letter code. Residues coded by IF-2_{chl} DNA are underlined.

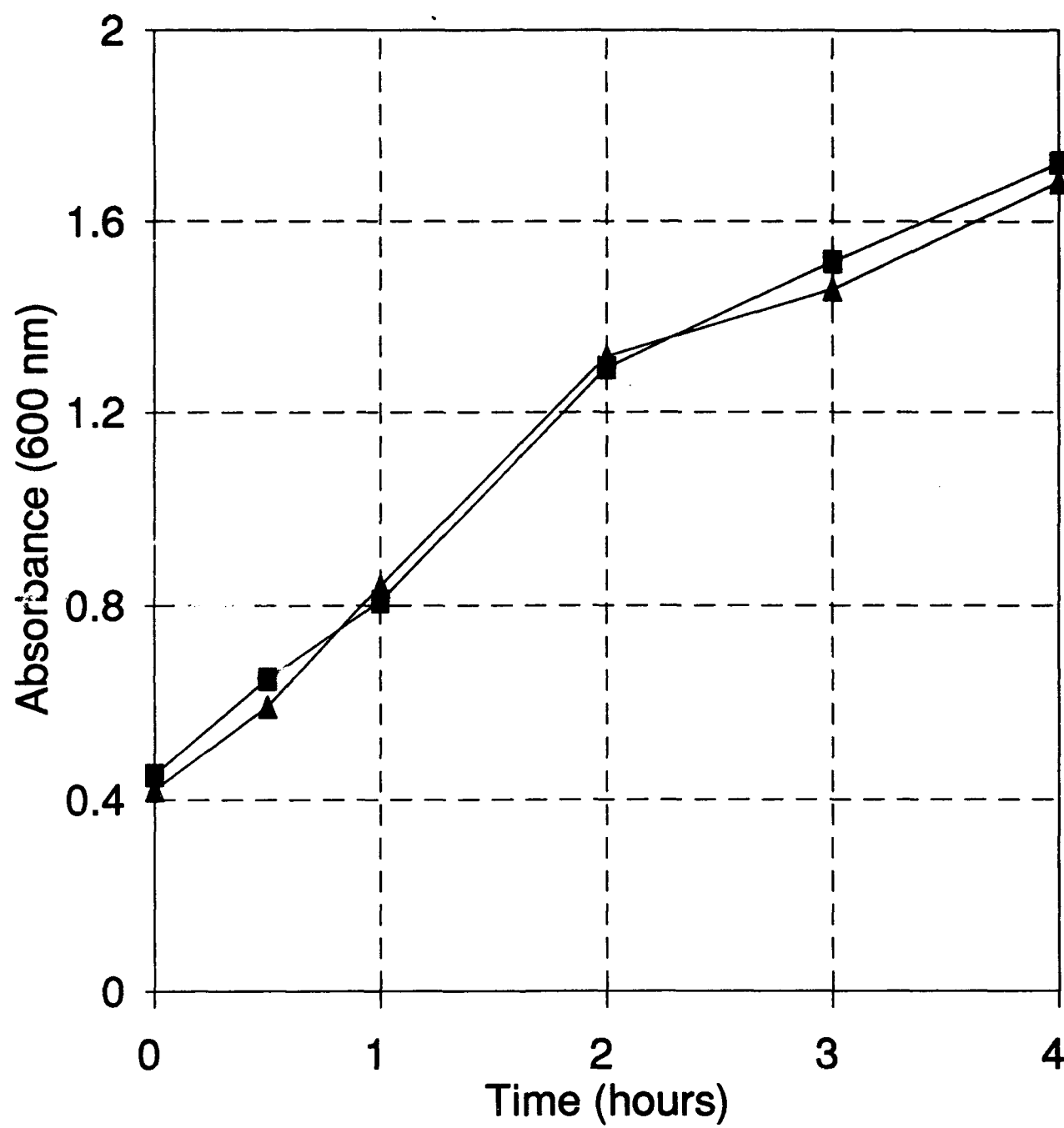
1 MRIPMRRRYL GMTVSEIAGK LAITPANVVT VLFKKGIMSA PSQTIAYDLV
51 KIVCDEYKVE VLEVVEEDGI ASMEDREVL D EEAEALVSRP PVVTIMGHVD
101 HGKTSLLDYI RKSNNVAGEA SGITQAIGAY HVEFASPTDG TPTFISFIDT
151 PGHEAFTAMR ARGATVTDIT IIVVAADDGV RPOTKEAIAH CKAAGVPMVV
201 AINKIDKDGA DPERVMNELA OAGLVPEEWG GEVPTVKISA KKGLGIKELL
251 EMILLTAEVA DLKANPAAPA EGTVIEAYLD RTRGPVATVL VONGTLRAGD
301 VVVTNATWGR VRAIMDEKGA MLEAAPPSLP VQVLGLDDVP AAGDKFEVYA
351 SEKEARDKVD EFERTKKEKN WASLASRDLV RLDNNADGKG LEVMNVILKT
401 DVSGSCEAIR AALDTLPOTK IELRLILASP GDITVSDVNL AASTGSIILG
451 FNVDTFSAAE ALIKNLGIKC MTFDVIYDLV DOMKAVMEGK LGDEQIPEKA
501 GAEVKAVFA ARNGKKAAGC LVVAGRLVAP AFIEVLRKKK ILFSGOLFOL
551 RRMKDNVSEV GTDTECGVTL DDFDDWQEGD RIVCYSTVTR ORALEATPAD
601 LRSHHHHHH*

residues which were added by the pQE vector, such as the start codon, residues encoded in the polylinker, and the 6X histidine tag. The calculated molecular masses for these polypeptides are 29.7 kilodaltons (kDa) for the G-domain and 65.7 kDa for the gamma protein.

Growth Rates of Strains Containing Expression Plasmids: The growth rates of *E. coli* carrying the expression constructs was monitored during small-scale expression experiments and provided interesting preliminary results. The G-domain protein had no effect on the growth rate of *E. coli* M15 as shown by Figure 9. In contrast, the growth rate of cells expressing the IF-2_{chl} gamma protein was drastically reduced (Figure 10).

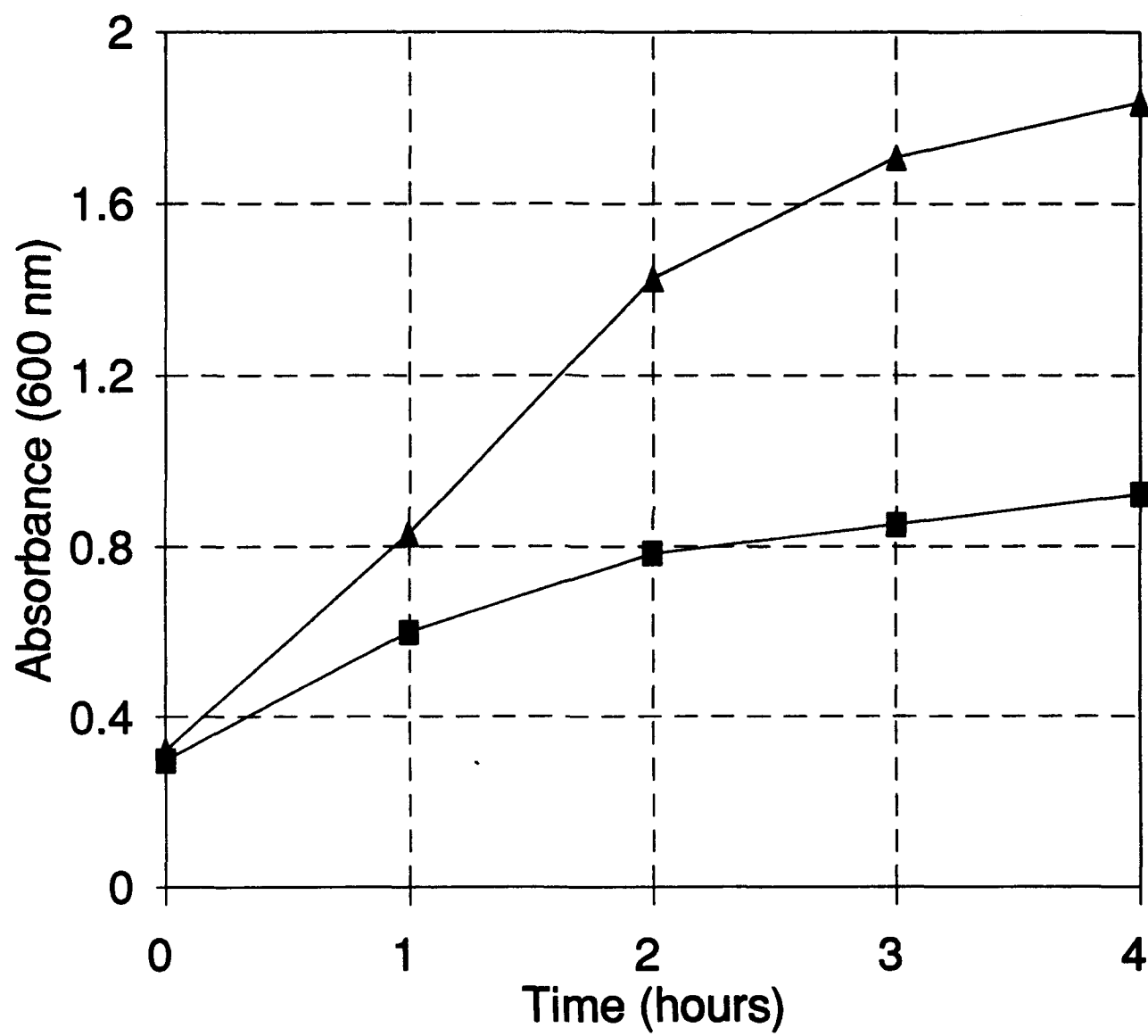
Small-Scale Expression of the G-domain Protein: Lysis under denaturing conditions and batch purification using Ni-NTA resin of the G-domain resulted in separation of a single protein from the cellular lysate (Figure 11). Only one protein bound the Ni-NTA resin, and that protein was not present prior to induction with IPTG at time = 0. The concentration of the protein slowly increased over time after induction. (Figure 11). The apparent size of the protein from SDS-PAGE is about 38 kDa which is relatively close to the calculated molecular mass of 29.7 kDa. Also, Qiong Lin, of this laboratory, observed smaller polypeptides

Figure 9: Growth Curve for *E. coli* M15 Carrying the G-Domain Construct. The plot illustrates the growth rate of *E. coli* M15 containing the G-domain construct, either induced with 1 mM IPTG or uninduced.



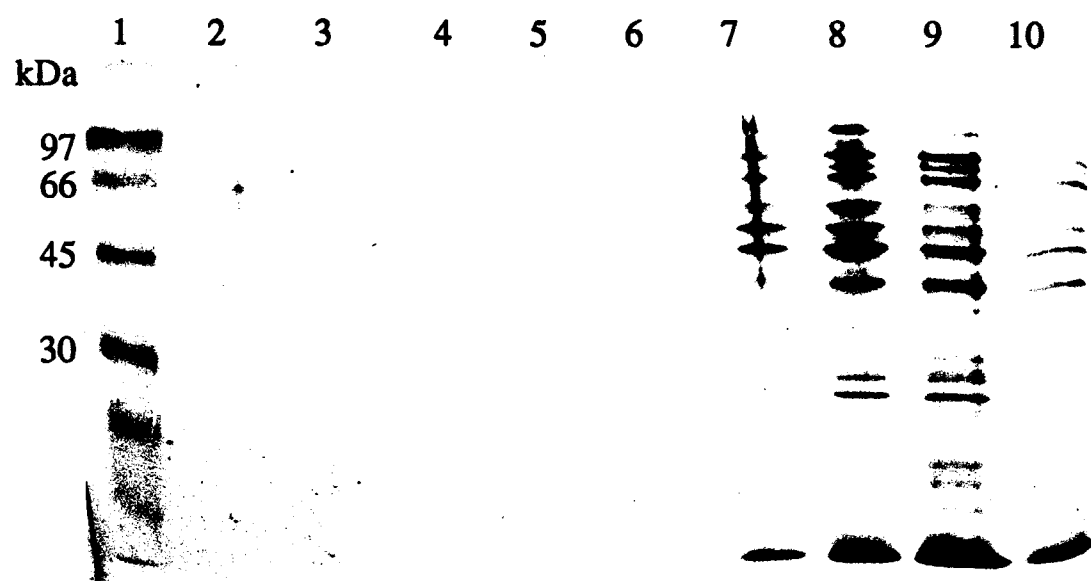
■ Induced ▲ Uninduced

Figure 10: Growth Curve for *E. coli* M15 Carrying the Gamma Construct. The plot illustrates the growth rate of *E. coli* M15 containing the gamma construct, either induced with 1 mM IPTG or uninduced.



■ Induced ▲ Uninduced

Figure 11: SDS-PAGE Analysis of the Small-Scale Expression of the G-Domain Protein. Lane 1 - Bio Rad pre-stained molecular weight standards. Lane 2 - Protein bound by Ni-NTA resin at time = 0 (prior to induction). Lanes 3-6 - Protein bound by Ni-NTA resin at time = 1, 2, 3, and 4 hours, respectively, after induction. Lanes 7-10 - Unbound protein at time = 0, 1, 3, and 4 hours, respectively.



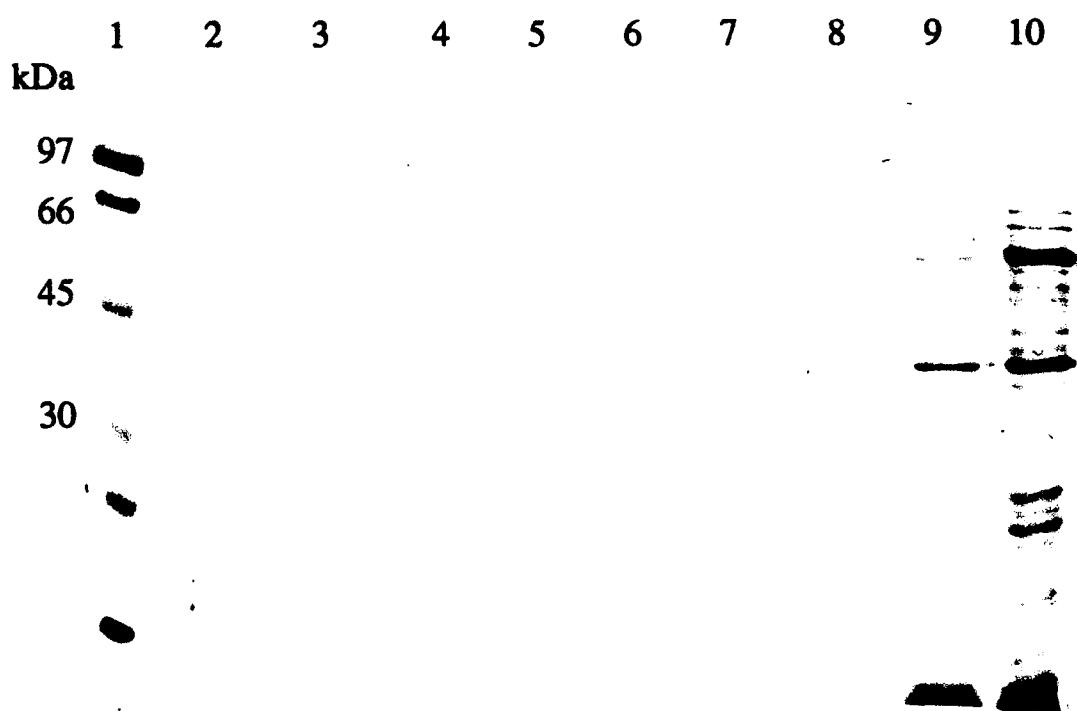
with the 6X His tag migrating slower than expected in SDS-PAGE experiments (personal communication).

Small-Scale Expression of the IF-2_{chl} Gamma Protein: Analysis of the Ni-NTA bound fraction from the lysate of the small scale expression of the gamma protein revealed a single protein with an apparent molecular mass of 65 kDa (Figure 12). This protein was not present at time = 0, prior to induction. A faint band can be seen as early as 15 minutes after induction (Figure 12). Unlike the G-domain protein, the relative concentration of IF-2_{chl} gamma decreased after 1 hour instead of steadily increasing (Figure 12). Analysis of the unbound proteins before and after induction shows the appearance of a 65 kDa protein after induction (Figure 12).

Purification of the G-Domain: The expression experiments described above were carried out on cell extracts prepared under denaturing conditions. In order to obtain the IF-2_{chl} G-domain in a native conformation, a purification scheme was developed. This four-step process started with thousands of contaminants and resulted in a preparation estimated to be 85-90 % pure.

The first step in the purification procedure, using the Ni-NTA affinity column removed many of the cellular proteins, but the G-domain preparation was clearly not pure.

Figure 12: SDS-PAGE Analysis of the Small-Scale Expression of IF-2_{chl} Gamma. Lane 1 - Bio Rad pre-stained molecular weight standards. Lane 2 - Proteins bound to Ni-NTA resin at time = 0 (prior to induction). Lanes 3-8 - Protein bound to Ni-NTA resin at times = 0.25, 0.5, 1, 2, 3, and 4 hours, respectively, after induction. Lane 9 - Unbound proteins at time = 0. Lane 10 - Unbound proteins at time = 0.5 hours.



(Figure 13). In Lane 2 (Figure 13) the large band at about 45 kDa is EF-Tu, and the somewhat bright band below it at 38-40 kDa is the G-domain. As is evident from the gel, the G-domain is greatly enriched in the eluted fractions (Lanes 5-8) and almost all of the EF-Tu, one of the most abundant proteins in *E. coli*, was separated away (Figure 13). However, hundreds of contaminants still remained after the Ni-NTA column (Figures 13 and 14). A majority of the G-domain eluted in the first 10 ml of the 20 ml gradient, and just small amounts could be seen after fraction 7 (Figures 13 and 14).

The first TSKgel DEAE-5PW HPLC column provided much better separation and approximately half of the contaminants present after the Ni-NTA column were removed (Figure 15). Retention of the G-domain by the DEAE-5PW column was excellent. About halfway through the gradient there was a strong absorbance at 280 nm (results not shown) which corresponded with the appearance of the G-domain on polyacrylamide gels (Figure 15). The first strong band for the G-domain (Figure 15, Lane 11) was in fraction 26 at an approximate NH_4Cl concentration of 175 mM.

The second DEAE-5PW HPLC column produced even better results than the first (Figures 16 and 17). The drawn out gradient and excellent retention produced very good separation leaving the G-domain as the major protein constituent in 4 fractions with as few as six unwanted

Figure 13: SDS-PAGE Analysis of Ni-NTA Column Fractions for the G-Domain. Lane 1 - Bio Rad pre-stained molecular weight standards. Lane 2 - Flow-thru from the Ni-NTA column. Lane 3 - Wash with pH 6.0 Buffer I. Lanes 4-10 - Imidazole eluted fractions 3-9, respectively. (The intense band with an apparent molecular mass of 38 kDa in lanes 3-8 is the G-domain protein.)

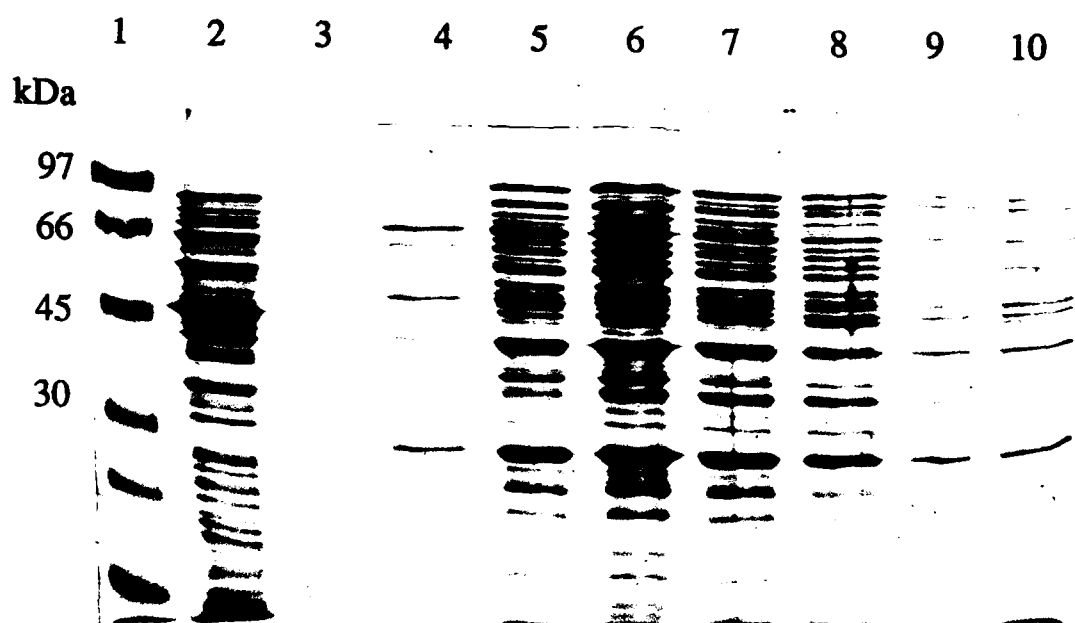


Figure 14: SDS-PAGE Analysis of Additional Ni-NTA Column Fractions for the G-Domain. Lane 1 - Bio Rad pre-stained molecular weight standards. Lanes 2-10 - Eluted fractions 11-18, respectively. Analysis of earlier column fractions is shown in Figure 13. (The G-domain protein is the protein with an apparent molecular mass of 38 kDa in lanes 2-8.)

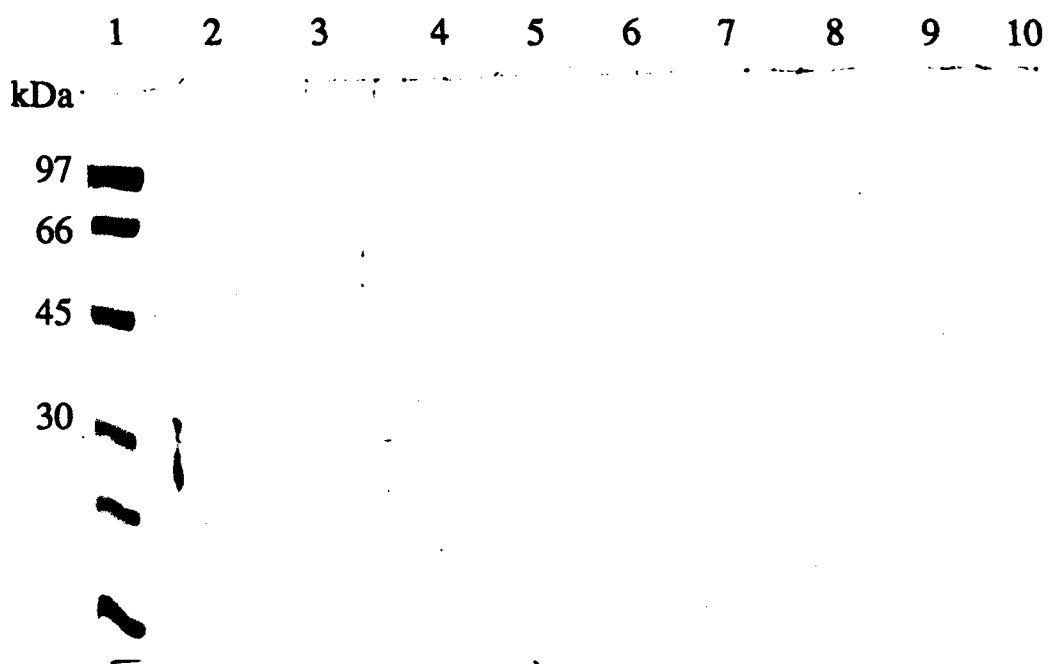
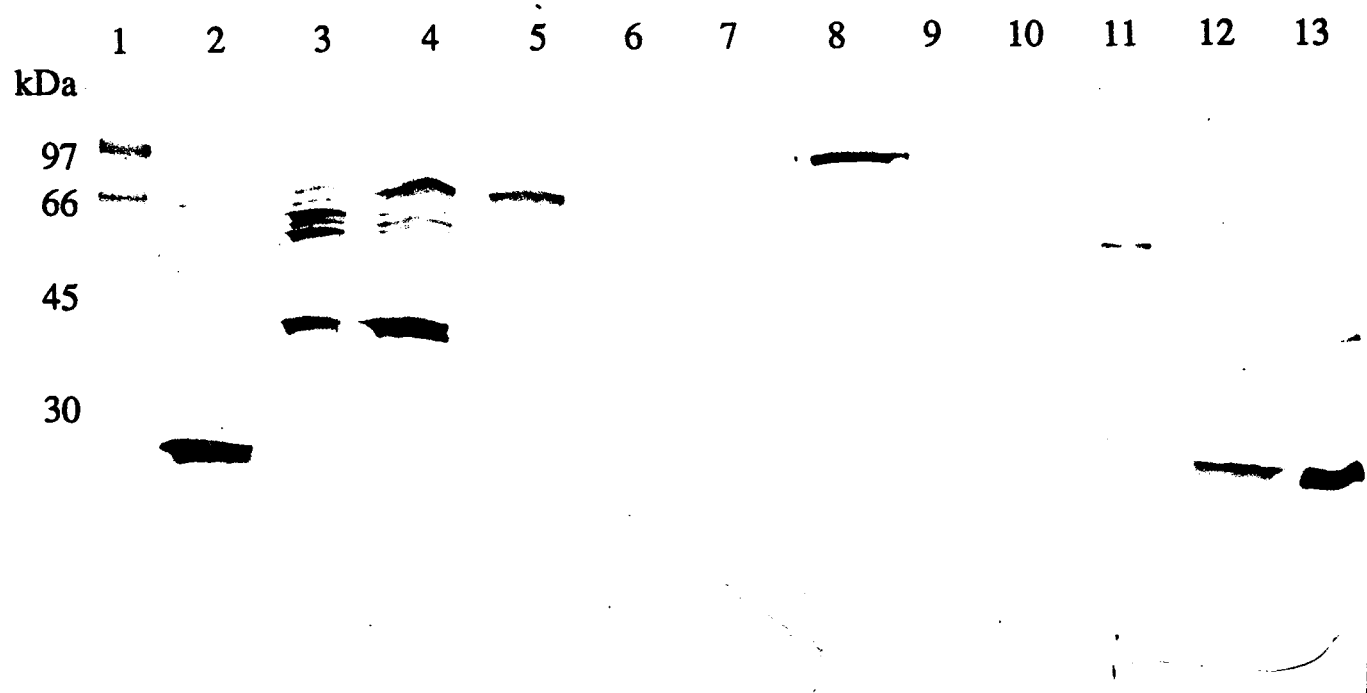


Figure 15: SDS-PAGE Analysis of Fractions from First TSKgel DEAE-5PW HPLC Column for the G-Domain. Lane 1 - Bio Rad pre-stained molecular weight standards. Lane 2-10 - Column fractions 29, 30, 31, 32, 34, 36, 37, 39, and 43, respectively. Lanes 11-13 - Column fractions 26-28.



**Figure 16: SDS-PAGE Analysis of Fractions from Second TSKgel
DEAE-5PW HPLC Column for the G-Domain. Lane 1 - Bio Rad
pre-stained molecular weight standards. Lanes 2-10 -
Fractions 74-82.**

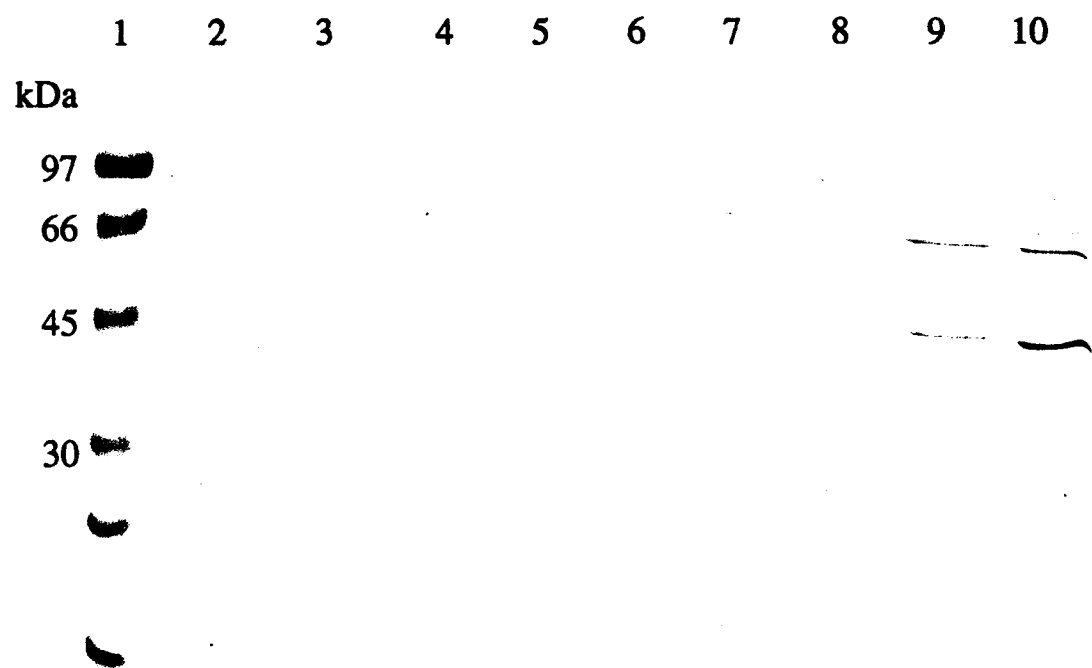


Figure 17: SDS-PAGE Analysis of Fractions from Second TSKgel DEAE-5PW HPLC Column for the G-Domain. Lane 1 - Bio Rad pre-stained molecular weight standards. Lanes 2-10 - Fractions 83-91. Analysis of earlier fractions in the gradient are shown in Figure 16.

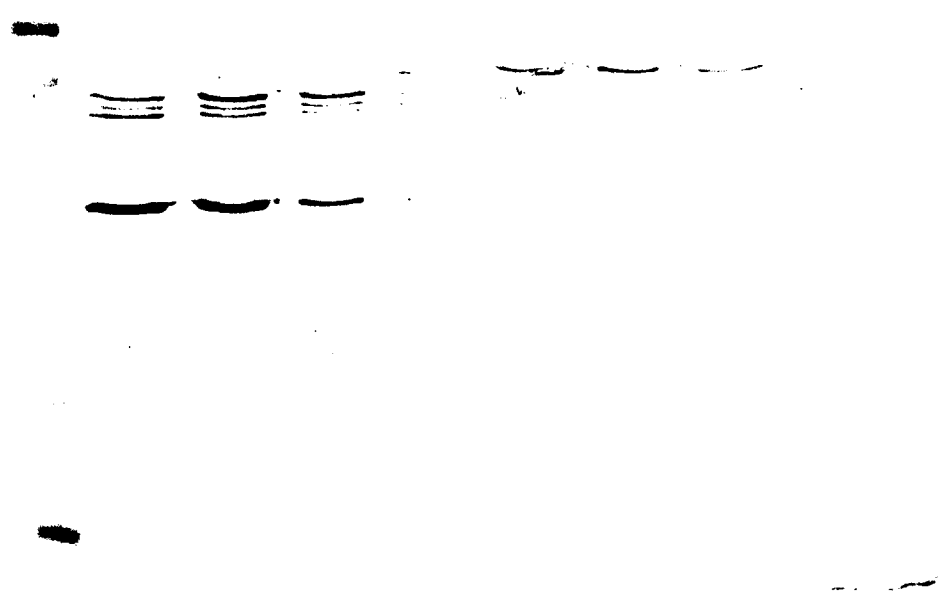
1 2 3 4 5 6 7 8 9 10
kDa

97

66

45

30



proteins present in lower amounts (Figures 16 and 17).

The TSKgel SP-5PW HPLC column provided little additional purification of the G-domain. Very little of the protein was retained by the column as indicated by the absorbance monitor and SDS-PAGE analysis of the flow-thru and fractions (results not shown). In fact, the bands on the gel were too faint to photograph and are not shown. The G-domain that was retained on the column appeared slightly more pure, with only four other proteins present. However, so little of the G-domain protein remained that the further purification of this particular preparation was abandoned.

DISCUSSION

Two truncated forms of IF-2_{chl} have been engineered at the DNA level and expressed in *E. coli*. These proteins, designated the G-domain and IF-2_{chl} gamma based upon similar work done with *E. coli* IF-2 (4,5), were expressed to different levels in *E. coli*.

Growth curves and relative amounts of protein on polyacrylamide gels indicated that the G-domain was produced and tolerated extremely well by *E. coli*. In contrast, the IF-2_{chl} gamma greatly retarded growth of *E. coli* and its relative amount in the cells peaked at about one hour. A possible explanation for this phenomena is that the gamma form possesses some IF-2 activity (either binding initiator tRNA or interacting with the ribosome) which is toxic to *E. coli*. Based on work in *E. coli*, the G-domain would not be expected to have any function except weak binding of GDP (4) which would have a limited effect on cell growth.

These two proteins were shown to have the 6X His tag, which is part of the expression vector, based on sequencing of the DNA and their affinity for Ni-NTA resin under denaturing conditions. Both proteins were bound by the Ni-NTA resin with high selectivity under denaturing conditions, probably due to the accessibility of the 6X His tag. However, under native conditions the G-domain was not bound as selectively.

Before an analysis of the purification of the G-domain can begin, the necessity of its purity must be discussed. The only IF-2 type activity found for the *E. coli* IF-2 G-domain was GDP binding that was much weaker than native IF-2 (4). Therefore, any trace contamination of *E. coli* IF-2 would drastically affect any binding assays attempted with the IF-2_{chl} G-domain. Due to this complication, the purification scheme developed had to be carried out without an assay for the factor, and relied only on the physical monitoring of the protein.

The Ni-NTA affinity column did not nearly duplicate the separation of the native G-domain protein in comparison to the denatured form. There are several explanations for this phenomena. As alluded to, the 6X His tag is probably more accessible in the denatured form, so the native form is bound less selectively. Also, the wash was ineffective due to the buffer system as illustrated by SDS-PAGE. Finally, some endogenous *E. coli* proteins may have adjacent histidines in their native state which are bound by the resin.

The TSKgel DEAE-5PW HPLC column provided excellent separation, except for those few proteins which were so similar electrostatically that they co-eluted. The use of the TSKgel SP-5PW HPLC column at pH 6.0 should have allowed evasion of this similarity based on a normal pKa for histidine between pH 6.0 and 7.0. However, upon later

computer analysis of the entire sequence a calculated pKa of 5.29 was discovered.

Based on experience gained in this investigation, the following changes to this scheme in future experiments would probably yield pure G-domain. Growth of a larger cell culture and a more stringent wash of the Ni-NTA affinity column would enhance the first step. Forego, the first DEAE HPLC column; and, assuming no solubility problems, lower the pH of the protein solution to 5.0 before applying it to the TSKgel SP-5PW HPLC column. These changes should enhance the purity.

The expression and purification described above on the G-domain and gamma form of IF-2_{chl} lays the foundation for a host of future experiments. The G-domain can be assayed for its ability to bind GDP, as well as its ability to hydrolyze GTP. When the entire sequence of IF-2_{chl} is cloned, a full length expression construct as well as truncated forms corresponding in size to *E. coli* IF-2 alpha and beta can be designed and expressed. *In vitro* assays of IF-2_{chl} activity can be performed with all the proteins to determine if IF-2_{chl} is active as a monomer, and if its relationship to *E. coli* IF-2 extends to its structure and function (i.e. is IF-2_{chl} gamma active like *E. coli* IF-2 gamma?).

The design, expression, and purification of two truncated forms of IF-2_{chl} is the first small step in understanding how the structure of chloroplast initiation

factor 2 from *E. gracilis* dictates its role in translational initiation.

REFERENCES

1. Darnell, J., Lodish, H., and Baltimore, D. (1990) in *Molecular Cell Biology* pp. 1066-1068, W.H. Freeman and Company, New York.
2. Ma, L., and Spremulli, L.L. (1990) *J. Biol. Chem.* **265**, 13560-13565.
3. Ma, L., and Spremulli, L.L., unpublished results.
4. Laalami, S., Sacerdot, C., Vachon, G., Mortensen, K., Sperling-Petersen, H.U., Cenatiempo, Y., and Grunberg-Manago, M. (1991) *Biochimie* **73**, 1557-1566.
5. Cenatiempo, Y., Deville, F., Dondon, J., Grunberg-Manago, M., Sacerdot, C., Hershey, J.W.B., Hansen, H.F., Petersen, H.U., Clark, B.F.C., Kjeldgaard, M., la Cour, T.F.M., Mortensen, K.K., and Nyborg, J. (1987) *Biochemistry* **26**, 5070-5076.
6. Laalami, S., Putzer, H., Plumbridge, J.A., and Grunberg-Manago, M. (1991) *J. Mol. Biol.* **220**, 335-349.
7. Lewin, B. (1990) in *Genes IV* pp. 122-124, Oxford University Press, Oxford.
8. Kraus, B. (1987) Identification and Characterization of a Chloroplast Ribosome Dissociation Factor, Doctoral Dissertation, Universtiy of North Carolina at Chapel Hill.
9. Ma, L., Ward, J., Blanchard, K., and Spremulli, L.L., unpublished results.

10. Crowe, J., and Henco, K. (1992) in *QIAexpress* pp. 24-43,
QIAGEN Incorporated, Chatsworth, California.
11. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) *Proc.*
Natl. Acad. Sci. USA **74**, 5463-5467.
12. Laemmli, U.K. (1970) *Nature* **227**, 680-685.

APPENDIX A

**Sequence Comparison Of IF-2 From *E. coli* And Chloroplasts
Of *E. gracilis***

Ecoif2	LA	AERQTSVE	L	VQQ	ADAGIR	KSADD	35
Egrif2chFQ	SSGSPIKP	I	NLD	PSTSTP	TPPEA	27
Ecoif2	SV	AQEK	TL	IDHLNQKNSG	PDKLTL	RKT	70
Egrif2ch	PT	PSAR	PV	TQVPQANSVP	AGAVAS	AEV	62
Ecoif2	IPGTGGK	KS	Q	VKRD	QEAE	LAAE	105
Egrif2ch	PQPPATP	AP	L.....R	RTAM	ASPP	MVIN	92
Ecoif2	EQAQREAE EQ	ARREAEES		REAQQ	AERE	AAE	140
Egrif2ch	LDDIPDRSKP	VWPAPPP		GQGGG	GGKG	GKG..	125
Ecoif2					T	NAQ	175
Egrif2ch					G	GGK	135
Ecoif2		E	LK		EA	V	210
Egrif2ch		P	..VV		KP	T	158
Ecoif2	E	N		PTEDSSD			245
Egrif2ch	A	S	SGGREAQ	P	175
Ecoif2		E	R	GR	NA	A	280
Egrif2ch	..K	K	DK	K	G	196
Ecoif2	AV	G	GKGGK	KGSS	LQQGFQ	KP	313
Egrif2ch	LL	L	KARKKT	E	KEHH	EEV	231
Ecoif2		ETI	G	L	N	M	347
Egrif2ch		QGM	S	I	G	L	266
Ecoif2	V	DQETAQL		AE	G	HK	381
Egrif2ch	T	AYDLVKI		CD	K	VE	300
Ecoif2		DTGA	AE	P	A		413
Egrif2ch		EEAE	LV	S	P		335
Ecoif2	TK	AS	G		H		442
Egrif2ch	SN	VA	S		A		370
Ecoif2	L		A	S		QA	477
Egrif2ch	I		E	A		TV	405
Ecoif2		I	Q	A		Q	512
Egrif2ch		K	A	C		G	440
Ecoif2		S	I	L		SQF	547
Egrif2ch		A	A	LV		VPT	475
Ecoif2		Q		LE		VRKGM	582
Egrif2ch		T		AD		NPAAP	510

Ecoif2	REH	Y	I	LCGFE	MRN	QEV	617
Egrif2ch	QNR	A	V	VTNAT	IMD	K AML	545

Ecoif2	LEAGI	I	SG	EVT	VRD	K E	652
Egrif2ch	EAPGL	Q	V	DD	AS	E D	580

Ecoif2	VALYRQG	R	VK	QOK	SK	ENMFAM	TEGEV	686
Egrif2ch	DEFERTK	K	KN	SLAS	D	VRLDN	DGKGL	614

Ecoif2	HEVIV	A	Q	V	SD	SLK	STDEV	KVKI	721
Egrif2ch	EVMVI	K	S	C	A	ADT	PQTKI	ELRL	649

Ecoif2	GSGVGG	ET	AT	ASNA	LV	RAD	AS	ARK	756
Egrif2ch	LASPGD	VS	DN	STGS	IL	DTF	SA	EA	684

Ecoif2	VEAESLDLR	YYS	N	I	EV	AS	SM	SP	LK	791
Egrif2ch	LKKNLG I K	M	TFD	D	V	QMAV	ESK	GD	Q I	719

Ecoif2	QQIIGL	VR	D	KSPKFGA	I	M	TE	V	VKRHN	826
Egrif2ch	PEKAGE	K	A	AARNGKK	A	L	V	AGR	LVAPA	754

Ecoif2	PIRVLRDNVV	IYE	EE	ES	ED	DN	NR	NGM	861	
Egrif2ch	FIEVLRKKKI	LFS	Q	FQ	KL	N	SE	VG	TDT	789

Ecoif2	GIGVKNYN	LV	RT	GEV	EVFE	IIEI	RTIA	...	890
Egrif2ch	GVTLD DFD	D	QEG	R	IVCYS	TVTR	ALEA	...	822

APPENDIX B

Known DNA And Translated Amino Acid Sequence Of IF-2_{chl}

1 TTCCAGTCCTCTGGCAGCCCTATCAAGCCCCGCATCAACCTTGACCGCCCCCTCCACCTCC 60
 F Q S S G S P I K P R I N L D R P S T S -
 61 ACCCCAACGCCCCCGGAAGCACCGACCAGCCCCCTCAGCCCGCCAGCCGGTGACGCAAGTG 120
 T P T P P E A P T S P S A R Q P V T Q V -
 121 CCCAGGCGAACAGCGTCCCTGCTGGAGCGGTAGCATCTCAAGCAGAAGTTAAGAAGCCG 180
 P Q A N S V P A G A V A S Q A E V K K P -
 181 GCAGACCCCCAGCCCCCGGCCACCCCCTCCGCGCCGGTGCTGCGGCGCCCCGTCCGCACC 240
 A D P Q P P A T P S A P V L R R P V R T -
 241 GCCATGCCGGCCTCGCCGCCCCGGATGGTGATCAACCTCGACGACATCCCCGACCGCTCC 300
 A M P A S P P R M V I N L D D I P D R S -
 301 AAGCCGGTGTGGCCCGCCCCCGCCCCGAGCAAAGGGGCAAGGGGGCGGCAAGGGCGGC 360
 K P V W P A P P P R A K G Q G G G K G G -
 361 AAAGGCGGCAAAGGCGGCAAAGGCGGCAAAGGTGGCAAGGGGGACCGCGAGCAGCCGGCG 420
 K G G K G G K G G K G G K G D R E Q P A -
 421 GTCGTGCGGCGGGCAAACACGGAGGACGGCAAGCACAGCCGAAGGGCCCCGCCGCGGAG 480
 V V R R A K P R R T A S T A E G P A A E -
 481 TCCAAAGAAAGCGGAGGACGCGAAGCCCAAATTTGGGTGACGCCCAAGGGTGGAAGGGC 540
 S K E S G G R E A Q I W V T P K G G K G -
 541 CGTGACAAGTGGAAGAAAGGAAAGGAGGAAGTCGACAAGAGCGAGGCGCTGCTGTTGAAA 600
 R D K W K K G K E E V D K S E A L L L K -
 601 GCCCGAAAGAAGACGCGCCTGGAGCGAAAGGAACGCCGCGAGGAGGTGCGGGAGGCGAAC 660
 A R K K T R L E R K E R R E E V R E A N -
 661 GCCGCCAAGAAGGAAGAGATTATCGAAGTTGGGCCGCAGGGGATGACCGTCAGCGAGATT 720
 A A K K E E I I E V G P Q G M T V S E I -
 721 GCCGGCAAACCTCGCCATCACCCCTGCCAACGTGGTGACAGTCCTCTTCAAGAAGGGCATC 780
 A G K L A I T P A N V V T V L F K K G I -

781 ATGAGTGGCCCGAGCCAGACGATTGCGTACGACCTCGTCAAGATTGTGTGCGATGAGTAC 840
 -----+-----+-----+-----+-----+-----+-----+
 M S A P S Q T I A Y D L V K I V C D E Y -
 841 AAAGTGGAAAGTTCTTGAAGTTGAGGAAGAAGACGGGATCGCGTCCATGGAGGACCGCTTC 900
 -----+-----+-----+-----+-----+-----+-----+
 K V E V L E V E E E D G I A S M E D R F -
 901 GTGCTGGACGAGGAGGCCGAAGCGTTGGTGTCCCGCCCGCGGTGGTGACCATCATGGGC 960
 -----+-----+-----+-----+-----+-----+-----+
 V L D E E A E A L V S R P P V V T I M G -
 961 CACGTGGACCACGGCAAGACCTCGCTGCTGGACTACATCCGCAAGTCCAACGTTGTCGCC 1020
 -----+-----+-----+-----+-----+-----+-----+
 H V D H G K T S L L D Y I R K S N V V A -
 1021 GGGGAGGCGAGCGGCATCACCCAGGCCATTGGCGCGTACCACGTCTGAATTCGCATCTCCC 1080
 -----+-----+-----+-----+-----+-----+-----+
 G E A S G I T Q A I G A Y H V E F A S P -
 1081 ACGGACGGCACCCTGACATTCATCTCATTTCATCGACACGCCGGGGCACGAGGCCTTCACG 1140
 -----+-----+-----+-----+-----+-----+-----+
 T D G T P T F I S F I D T P G H E A F T -
 1141 GCGATGCGGGCCCGCGGGGCTACCGTGACGGATATCACCATCATCGTGGTGGCAGCAGAC 1200
 -----+-----+-----+-----+-----+-----+-----+
 A M R A R G A T V T D I T I I V V A A D -
 1201 GACGGGGTGCGGCCCCAGACCAAGGAGGCTATTGCTCACTGCAAGGCCGCTGGGGTGCCA 1260
 -----+-----+-----+-----+-----+-----+-----+
 D G V R P Q T K E A I A H C K A A G V P -
 1261 ATGGTGGTGGCGATCAACAAGATCGACAAGGACGGCGCGGACCCGGAGCGGGTGATGAAC 1320
 -----+-----+-----+-----+-----+-----+-----+
 M V V A I N K I D K D G A D P E R V M N -
 1321 GAGCTGGCGCAGGCGGGGCTGGTGCCGAGGAGTGGGGCGGCGAGGTGCCGACGGTGAAG 1380
 -----+-----+-----+-----+-----+-----+-----+
 E L A Q A G L V P E E W G G E V P T V K -
 1381 ATCAGCGCCAAGAAGGGCCTCGGCATCAAGGAGCTGCTGGAGATGATCCTCCTCACCGCG 1440
 -----+-----+-----+-----+-----+-----+-----+
 I S A K K G L G I K E L L E M I L L T A -
 1441 GAGGTGGCGGACCTGAAGGCCAACCCCGCGGCCCCCGCGGAGGGCACTGTCATTGAGGCA 1500
 -----+-----+-----+-----+-----+-----+-----+
 E V A D L K A N P A A P A E G T V I E A -
 1501 TATTTGGACCGAACACGCGGGCCGGTTGCGACGGTGCTCGTCCAAAACGGCACTCTGCGG 1560
 -----+-----+-----+-----+-----+-----+-----+
 Y L D R T R G P V A T V L V Q N G T L R -

1561 GCGGGCGACGTGGTAGTCACCAACGCTACCTGGGGCCGGGTGCGGGCCATCATGGACGAG 1620
 -----+-----+-----+-----+-----+-----+-----+
 A G D V V V T N A T W G R V R A I M D E -
 1621 AAGGGGGCAATGCTGGAGGCTGCGCCCCGTCGCTGCCCGTCCAAGTGCTCGGCCTGGAC 1680
 -----+-----+-----+-----+-----+-----+-----+
 K G A M L E A A P P S L P V Q V L G L D -
 1681 GACGTCCCAGCCGCTGGGGACAAGTTTGAGGTCTACGCGTCGGAGAAGGAGGCGAGGGAC 1740
 -----+-----+-----+-----+-----+-----+-----+
 D V P A A G D K F E V Y A S E K E A R D -
 1741 AAGGTGGACGAGTTTGAGCGGACCAAGAAGGAGAAGAATTGGGCGTCGCTGGCGTCCCGG 1800
 -----+-----+-----+-----+-----+-----+-----+
 K V D E F E R T K K E K N W A S L A S R -
 1801 GACTTGGTGCGGCTGGACAACAACGCGGATGGGAAGGGGTGGAGGTGATGAACGTCATC 1860
 -----+-----+-----+-----+-----+-----+-----+
 D L V R L D N N A D G K G L E V M N V I -
 1861 CTCAAGACCGACGTCTCCGGGTCCTGCGAGGCCATCCGGGCGGGCGCTGGACACCCTGCCC 1920
 -----+-----+-----+-----+-----+-----+-----+
 L K T D V S G S C E A I R A A L D T L P -
 1921 CAGACCAAGATCGAGCTGCGCCTGATCCTGGCCTCCCCGGGGGACATCACCGTCTCCGAT 1980
 -----+-----+-----+-----+-----+-----+-----+
 Q T K I E L R L I L A S P G D I T V S D -
 1981 GTCAACCTTGCTGCTTCCACGGGTAGCATCATCCTGGGCTTCAACGTGGACACGTTTTCT 2040
 -----+-----+-----+-----+-----+-----+-----+
 V N L A A S T G S I I L G F N V D T F S -
 2041 GCCGCGGAGGCGCTCATCAAGAACCTTGGCATCAAGTGCATGACGTTTGATGTCATCTAC 2100
 -----+-----+-----+-----+-----+-----+-----+
 A A E A L I K N L G I K C M T F D V I Y -
 2101 GACCTCGTGGACCAGATGAAGGCAGTAATGGAAGGCAAGTTGGGCGACGAGCAGATCCCG 2160
 -----+-----+-----+-----+-----+-----+-----+
 D L V D Q M K A V M E G K L G D E Q I P -
 2161 GAGAAGGCCGGGGAGGCGGAGGTGAAGGCGGTGTTCGCGGCGCGGAACGGGAAGAAGGCA 2220
 -----+-----+-----+-----+-----+-----+-----+
 E K A G E A E V K A V F A A R N G K K A -
 2221 GCCGGGTGCCTGGTGGTCGCGGGCCGCTGGTCGCCCCGCGTTCATCGAGGTGCTGCGG 2280
 -----+-----+-----+-----+-----+-----+-----+
 A G C L V V A G R L V A P A F I E V L R -
 2281 AAGAAGAAGATCCTGTTTTTCGGGGCAGCTGTTCCAAGTGCGGCGGATGAAGGACAATGTC 2340
 -----+-----+-----+-----+-----+-----+-----+
 K K K I L F S G Q L F Q L R R M K D N V -

2341 AGCGAGGTGGGCACCGACACCGAGTGCGGTGTCAACCTTGACGACTTCGACGACTGGCAG 2400
 -----+-----+-----+-----+-----+-----+-----+
 S E V G T D T E C G V T L D D F D D W Q -

 2401 GAGGGCGACCGCATCGTGTGCTACAGCACGGTCACCCGGCAACGGGCCCTCGAGGCGACC 2460
 -----+-----+-----+-----+-----+-----+-----+
 E G D R I V C Y S T V T R Q R A L E A T -

 CCGCT
 2461 ----- 2466
 P A -